Pergamon

Appl. Math. Lett. Vol. 11, No. 1, pp. 47–51, 1998

0893-9659/98 $19.00 + 0.00

PII: S0893-9659(97)00131-6

Finite and Continuous Perturbations of Matrix Eigenvalues

A. L. ANDREW
Department of Mathematics, La Trobe University
Bundoora, Victoria 3083, Australia

(Received April 1997; accepted May 1997)

Communicated by R. Tewarson

Abstract—An elementary proof is given that some well-known formulae for derivatives of eigenvalues of matrix-valued functions hold under weaker hypotheses than are required by the usual proofs. The relationship between continuous and finite perturbations is also discussed.

Keywords—Eigenvalue sensitivities, Multiple eigenvalues, Generalized eigenvalue problems.

1. INTRODUCTION

This letter considers two perturbation problems for the eigenvalue problem $Ax = \lambda Bx$, where A and B are $n \times n$ matrices. One is the finite perturbation problem, in which it is required to find bounds for the perturbation in the eigenvalues and eigenvectors produced by finite changes in the matrix elements, given bounds for the changes in A and B. This problem, which arises when there is uncertainty in the data and in assessing the effect of roundoff errors, is treated in detail in [1]. The other is the continuous perturbation problem

$$A(t)x(t) = \lambda(t)B(t)x(t), \qquad y^\top(t)A(t) = \lambda(t)y^\top(t)B(t), \tag{1}$$

where the matrices $A(t)$ and $B(t)$ and the eigenvalues and eigenvectors are continuous functions of a parameter t. Often these functions are differentiable, and in a number of problems in engineering, for example, the optimal design of structures [2] and model updating [3], it is useful to know the derivatives of the eigenvalues and, if possible, the eigenvectors. Often these derivatives are also used to estimate the effect of finite perturbations, although this can lead to underestimates of errors in computed solutions. Recall that $f : \mathbb{R}^m \to \mathbb{C}^s$ is said to be differentiable at $t \in \mathbb{R}^m$ if there exists a (complex) $s \times m$ matrix, $f'(t)$, such that $f(t+\delta) - f(t) = f'(t)\delta + o(\delta)$ as $\delta \to 0$. The prime $'$ is used in this letter to denote derivatives defined in this sense.

Usually finite and continuous perturbations are studied using different techniques. This letter adopts a more unified approach by using a common starting point, equation (4) below, to study both. An important advantage of this approach is that it establishes sufficient conditions for the differentiability of the eigenvalues which are weaker than the classical conditions, at least in the case of repeated eigenvalues. In particular, the eigenvectors are not required to be differentiable. In Section 3 these results are extended to the more general eigenvalue problem

$$L(t,\lambda(t))\,x(t) = 0, \qquad y^\top(t)L(t,\lambda(t)) = 0^\top, \tag{2}$$

Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX where L is a differentiable matrix-valued function. This problem was studied in some detail under slightly more restrictive smoothness conditions in [4], where references are also given for some applications of such more general problems.

2. TOWARDS A UNIFIED APPROACH

Let $x_i(t), y_i(t)$ be right and left eigenvectors, respectively, of (1) corresponding to an eigenvalue $\lambda_i(t)$ at the point $t \in \mathbb{R}$. The formula

$$\lambda_i' = \frac{y_i^\top (A' - \lambda_i B') x_i}{y_i^\top B x_i} \tag{3}$$

is usually established by *assuming* that λ_i and x_i are differentiable, and differentiating (1) to obtain $A'x_i + Ax_i' = \lambda_i' Bx_i + \lambda_i B' x_i + \lambda_i B x_i'$ from which (3) follows on premultiplying by $y_i^\top$, whenever $(y_i^\top B x_i)(t) \neq 0$. Another common approach, used in [5,6], for example, is to consider only the case in which λ_i and x_i are analytic and expand the terms in (1) in a power series. Essentially this approach was used by Jacobi, who considered only first-order terms and derived a result [7, equation (12)] which, when translated into modern notation, becomes the special case of (3) in which $B = I$ (the identity) and A is real symmetric.

It does not seem to be widely realised that there is an elementary derivation of (3) which makes no assumption about the differentiability of λ_i or x_i, and which helps clarify the relationship between continuous and finite perturbations. The basic observation is that (1) implies that, for $i, j = 1, \dots, n$ and for all $t, \delta \in \mathbb{R}$,

$$\begin{aligned} &y_j^\top(t)\left[A(t+\delta)x_i(t+\delta) - \lambda_i(t+\delta)B(t+\delta)x_i(t+\delta)\right] \\ &\quad = 0 = \left[y_j^\top(t)A(t) - \lambda_j(t)y_j^\top(t)B(t)\right] x_i(t+\delta), \end{aligned}$$

and hence,

$$\begin{aligned} &y_j^\top(t)B(t+\delta)x_i(t+\delta)\left[\lambda_i(t+\delta) - \lambda_j(t)\right] \\ &\quad = y_j^\top(t)\left\{\left[A(t+\delta) - A(t)\right] - \lambda_j(t)\left[B(t+\delta) - B(t)\right]\right\} x_i(t+\delta). \end{aligned} \tag{4}$$

Putting $j = i$ in (4), dividing both sides by δ, and taking the limit as $\delta \to 0$ shows that the following three conditions are sufficient for (3) to hold at any point t:

(i) A and B are differentiable at t;
(ii) $y_i^\top(t)B(t)x_i(t) \neq 0$; and
(iii) x_i is continuous at t.

The first two are also necessary for the right-hand side of (3) to exist at t.

Two examples in which the eigenvectors are continuous everywhere, but not differentiable when $t = 0$, are given by

$$A(t) = \begin{bmatrix} t^3 & t|t| \\ t|t| & t \end{bmatrix}, \qquad \text{and by } A(t) = \begin{bmatrix} 0 & t \\ t^2 & 0 \end{bmatrix},$$

with $B(t) = I$ in both cases. For all $t \in \mathbb{R}$, the first example satisfies all the above three sufficient conditions, and hence, the above proof shows that (3) gives the correct value of the eigenvalue derivatives, as is also readily checked directly. The continuous eigenvectors of the second example do not satisfy $(y_i^\top B x_i)(0) \neq 0$, but all three sufficient conditions are satisfied at all $t \neq 0$, and since the eigenvalue derivatives are continuous for that example, the correct eigenvalue derivatives at $t = 0$ are given by the limit at $t = 0$ of the right-hand side of (3).

Although (3) and related results are derived in numerous papers and books, none of those known to the author mentioned the ease with which (3) may be derived, under weaker assumptions than those traditionally made, from the case $i = j$ of (4). However, another special case of (4), that in which A and B are constant, is used in the standard proof that (for fixed t) $y_j^\top B x_i = 0$ whenever

$\lambda_i \neq \lambda_j$, and this in turn is used to show that, corresponding to each semisimple eigenvalue λ_i of (1), there is a left eigenvector y_i and a right eigenvector x_i such that $y_i^\top B x_i \neq 0$, as required for the validity of (3). (An eigenvalue is called "semisimple" if its multiplicity as a zero of the characteristic equation equals the dimension of the corresponding eigenspace.)

Although (3) is often used to estimate the effect of finite perturbations in $A(t)$ and $B(t)$ on eigenvalues of (1), it gives only a first order approximation for this, whereas (4) is an exact result. When $B = I$, for example, $\epsilon\|y_i\|_2\|x_i\|_2/|y_i^\top x_i|$ is often taken as a rough upper bound for the effect of a small perturbation with 2-norm ϵ on the simple eigenvalues of A (see [5, p. 69]), whereas (4) shows that a more reliable bound is

$$\sup_{\|\Delta\|_2 \leq \epsilon} \frac{\epsilon\|y_i\|_2\|x_i(\Delta)\|_2}{|y_i^\top x_i(\Delta)|},$$

where $x_i(\Delta)$ is the *appropriate* eigenvector of $A + \Delta$. When $A = \begin{bmatrix} 0 & 1 \\ \epsilon & 0 \end{bmatrix}$, this gives the bound $\epsilon^{1/2}(1+\epsilon)^{1/2}$, which is marginally greater than the exact eigenvalue perturbation $\epsilon^{1/2}$ obtained by a perturbation of $-\epsilon$ in the bottom left element of A, whereas the classical estimate underestimates the perturbation by a factor of approximately two.

It is well known that rigorous bounds which apply in *all* cases are generally much more difficult to compute than estimates or rough "bounds" which apply in "most" cases. The discussion in the previous paragraph is useful in highlighting the limitations of first order analysis of finite perturbations, but, while it helps clarify the difficulty of computing bounds for perturbations, it does not make computing such bounds easy. Even if Δ were known exactly, it may not be obvious which right eigenvector of $A + \Delta$ "corresponds" to the left eigenvector y_i of A. The choice will be much easier if there is a unique right eigenvector of $A + \Delta$ "close" to the right eigenvector of A corresponding to the same eigenvalue as y_i. Thus, the idea of continuity of eigenvectors also plays a role in the analysis of finite eigenvalue perturbation.

The assumption of continuity of eigenvectors is not trivial, however. A famous example of Rellich [6, p. 405], with $B = I$ and A real symmetric, and with both A and all eigenvalues of (1) infinitely differentiable on $\mathbb{R}$, has no eigenvectors continuous on $\mathbb{R}$. See, also, the discussion in [4].

Like the two examples given earlier with continuous eigenvectors not differentiable at $t = 0$, Rellich's example has a multiple eigenvalue. This is no coincidence. Eigenvectors corresponding to multiple eigenvalues are often less well behaved than those corresponding to simple eigenvalues. Another well known difficulty is that care is needed to choose the appropriate y_i and x_i in (3) when the eigenspace is multidimensional. Discussions of this in the literature (see [8,9] for references) generally consider only the case in which eigenvectors are differentiable and the criterion given is that $y_i(t)$ and $x_i(t)$ be chosen to ensure that the eigenvectors are differentiable. However, the choice is uniquely determined by the weaker requirement that the eigenvectors be continuous and the usual techniques remain valid when the eigenvectors are merely continuous.

For notational convenience, let λ_1 be a semisimple eigenvalue of (1) of multiplicity r at t and let $\lambda_1(t) = \cdots = \lambda_r(t)$. Let X and Y be $n \times r$ matrices whose columns are linearly independent continuous right and left eigenvectors, respectively, of (1) corresponding to $\lambda_1, \ldots, \lambda_r$, normalized by $Y^\top B X = I$, and let $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$. Then,

$$\begin{aligned} &Y^\top(t)B(t+\delta)X(t+\delta)\,[\Lambda(t+\delta) - \Lambda(t)] \\ &\quad = Y^\top(t)\,\{[A(t+\delta) - A(t)] - \lambda_1(t)\,[B(t+\delta) - B(t)]\}\,X(t+\delta). \end{aligned}$$

Then, as in the above derivation of (3), $\Lambda'(t) = [Y^\top(A' - \lambda_1 B')X](t)$. Let $\hat{X}$ and $\hat{Y}$ be $n \times r$ matrices whose columns are any bases of the right and left eigenspaces, respectively, corresponding to $\lambda_1(t)$, normalized by $\hat{Y}^\top B(t)\hat{X} = I$, obtained by some numerical computation. Then, for some nonsingular $r \times r$ matrix C, $\hat{X} = X(t)C$ and $\hat{Y}C^\top = Y(t)$. Substituting in the above equation for $\Lambda'(t)$ shows that the values of the derivatives $\lambda_1'(t), \ldots, \lambda_r'(t)$ are given by the eigenvalues of

$\hat{Y}^{\mathsf{T}}(A'(t)-\lambda_1(t)B'(t))\hat{X}$. Unlike the usual derivation of this result, the above argument does not require the eigenvectors to be differentiable. For the symmetric case with $B = I$, a useful but much less elementary treatment of eigenvalue derivatives with minimal smoothness conditions is given in [10].

3. SOME GENERALIZATIONS

The above derivation of (3) is readily extended to functions of several variables when, in (1), $t = (t_1, \ldots, t_m) \in \mathbb{R}^m$. The same argument shows that (4) holds for all $t, \delta \in \mathbb{R}^m$ and, since $[1+o(1)]^{-1} = 1 + o(1)$, the same three assumptions imply that for all sufficiently small δ,

$$\begin{aligned}\lambda_i(t+\delta)-\lambda_i(t) &= \left[y_i^{\mathsf{T}}(t)B(t+\delta)x_i(t+\delta)\right]^{-1} y_i^{\mathsf{T}}(t)\left\{A'(t)\delta \right.\\ &\qquad \left. + o(\delta) - \lambda_i(t)\left[B'(t)\delta + o(\delta)\right]\right\} x_i(t+\delta)\\ &= \sum_{k=1}^{m} \lambda_{i,k}(t)\delta_k + o(\delta),\end{aligned}$$

where $\delta = (\delta_1, \ldots, \delta_m)$, and for $k = 1, \ldots, m$,

$$\lambda_{i,k}(t) = \frac{y_i^{\mathsf{T}}(t)\left[A_{,k}(t) - \lambda_i(t)B_{,k}(t)\right]x_i(t)}{y_i^{\mathsf{T}}(t)B(t)x_i(t)}, \tag{5}$$

where $A_{,k}$ and $B_{,k}$ denote the partial derivatives of A and B, respectively, with respect to δ_k. It follows that λ_i is differentiable at t, and that its partial derivative with respect to δ_k is given by (5).

The more general problem (2) with $t = (t_1, \ldots, t_m) \in \mathbb{R}^m$ may be treated similarly. For $k = 1, \ldots, m$, let $L_k(t, \lambda)$ denote the partial derivative of $L(t, \lambda)$ with respect to t_k and $L^{(1)}(t, \lambda)$ the partial derivative with respect to λ, as in [4]. To prove existence of $\lambda_i'(t)$ for this more general problem and to determine its value, we make the following assumptions:

(i) L is differentiable at $(t, \lambda_i(t))$;
(ii) $y_i^{\mathsf{T}}(t)L^{(1)}(t, \lambda_i(t))x_i(t) \neq 0$;
(iii) the eigenvector x_i is continuous at t; and
(iv) $\lambda_i(t+\delta) - \lambda_i(t) = O(\delta)$ as $\delta \to 0$.

In the special case (1), the first three conditions reduce to the three conditions already given for (1). The additional fourth condition is weaker than differentiability of λ_i at t.

Since, by (2),

$$y_i^{\mathsf{T}}(t)L\left(t+\delta, \lambda_i(t+\delta)\right)x_i(t+\delta) = 0 = y_i^{\mathsf{T}}(t)L\left(t, \lambda_i(t)\right)x_i(t+\delta),$$

it follows from the first of the above four assumptions that

$$\begin{aligned} y_i^{\mathsf{T}}(t)\Bigg[\sum_{k=1}^{m} L_k\left(t, \lambda_i(t)\right)\delta_k + L^{(1)}\left(t, \lambda_i(t)\right)\left(\lambda_i(t+\delta) - \lambda_i(t)\right) + o(\delta)\\ + o\left(\lambda_i(t+\delta) - \lambda_i(t)\right)\Bigg] x_i(t+\delta) = 0,\end{aligned}$$

and hence, by a straightforward calculation using the other three assumptions,

$$\lambda_i(t+\delta) - \lambda_i(t) = \sum_{k=1}^{m} \lambda_{i,k}(t)\delta_k + o(\delta),$$

where again $\delta = (\delta_1, \ldots, \delta_m)$ but this time, for $k = 1, \ldots, m$,

$$\lambda_{i,k}(t) = \frac{-y_i^{\mathsf{T}}(t)L_k\left(t, \lambda_i(t)\right)x_i(t)}{\left[y_i^{\mathsf{T}}(t)L^{(1)}\left(t, \lambda_i(t)\right)x_i(t)\right]}. \tag{6}$$

Hence, λ_i is differentiable at t and its partial derivative with respect to δ_k at t is given by (6). Note, this formula reduces to (5) when $L(t, \lambda(t)) = A(t) - \lambda(t)B(t)$. The formula (6) is well known, but all previous derivations known to the author have assumed that both λ_i and x_i are differentiable at t.

REFERENCES

1. G.W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*, Academic Press, New York, (1990).
2. E.J. Haug, K.K. Choi and V. Komkov, *Design Sensitivity Analysis of Structural Systems*, Academic Press, New York, (1986).
3. J.E. Mottershead and M.I. Friswell, Model updating in structural dynamics: A survey, *J. Sound Vibration* **167**, 347–375, (1993).
4. A.L. Andrew, K.-W.E. Chu and P. Lancaster, Derivatives of eigenvalues and eigenvectors of matrix functions, *SIAM J. Matrix Anal. Appl.* **14**, 903–926, (1993).
5. J.H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, (1965).
6. P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, Academic Press, New York, (1985).
7. C.G.J. Jacobi, An easy method to solve numerically the equations occurring in the theory of secular perturbations, (English transl. NASA TT F-13,666 (1971)), *Z. Reine Angew. Math.* **30**, 51–95, (1846).
8. A.L. Andrew and R.C.E. Tan, Computation of derivatives of repeated eigenvalues and the corresponding eigenvectors of symmetric matrix pencils, Math Res. Paper 96-14, La Trobe University, Melbourne, (1996).
9. A.L. Andrew and R.C.E. Tan, Computation of derivatives of repeated eigenvalues and corresponding eigenvectors by simultaneous iteration, *AIAA J.* **34**, 2214–2216, (1996).
10. J.-B. Hiriart-Urruty and D. Ye, Sensitivity analysis of all eigenvalues of a symmetric matrix, *Numer. Math.* **70**, 45–72, (1995).